DOCUMENT RESUME

ED 449 219 TM 032 364

AUTHOR	Chow, Alan Ping-yan; Wong, Edwin King-por; Yeung, Alexander Sheesing; Mo, Kim Wan
TITLE	Appraisal of Teachers: Who Appraises Whom and Does It Matter?
PUB DATE	2001-01-00
NOTE	18p.; Paper presented at the International Congress for School Effectiveness and Improvement (14th, Toronto, Ontario, Canada, January 5-9, 2001).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Education; *Elementary School Teachers; Factor Analysis; Foreign Countries; Formative Evaluation; *Principals; Summative Evaluation; Teacher Administrator Relationship; Teacher Attitudes; *Teacher Evaluation
IDENTIFIERS	Hong Kong

ABSTRACT

Elementary school teachers in Hong Kong (n=527) responded to survey items about formative outcomes, summative outcomes, perceived purposes of appraisal, overall effectiveness of appraisal, and summative purposes such as promotion and dismissal of staff. Principal components analysis and confirmatory analysis yielded the two a priori outcome factors, each of which was significantly correlated with perceived overall effectiveness of appraisal. When analyzed separately, analysis of variance found that senior teachers appraised by the school principal (SP group) had a stronger sense of the formative purpose of appraisal than teachers appraised by senior staff (TS group), while the TS group had the weakest sense of the promotion purposes. Teachers appraised by the principal (TP group) had a strong perception of the dismissal purpose than the TS group. The TS group was more favorable toward the competence of their appraisers than the TP group. Although the three groups did not differ in their perceptions of formative outcomes, summative outcomes, or the overall effectiveness of teacher appraisal, the appraiser-appraisee combination did make a significant difference in teacher perceptions of the purposes and appropriateness of the appraisal. An appendix contains a list of the survey items. (Contains 2 figures, 3 tables, and 33 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Paper Presentation

Appraisal of Teachers: Who Appraises Whom and Does it Matter?

Alan Ping-yan Chow, Edwin King-por Wong, Alexander Seeshing Yeung & Kim Wan Mo
Hong Kong Institute of Education

ED 449 219

Paper presented at
The 14th International Congress for School Effectiveness and Improvement
Toronto, Canada
5-9 January, 2001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
A. P.-Y. Chow
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

TM032364

Author Note

Correspondence concerning this paper should be sent to Alan Ping-yan Chow, Division of Continuing Professional Education, The Hong Kong Institute of Education, 10 Lo Ping Road, Tai Po, N.T., Hong Kong or via e-mail to pyachow@ied.edu.hk.

BEST COPY AVAILABLE

Appraisal of Teachers: Who Appraises Whom and Does it Matter?

Abstract

Teacher appraisal procedures may lead to formative (teacher development and improvement of teaching) and summative (managerial decision) outcomes. Elementary school teachers in Hong Kong (N = 527) responded to survey items on formative outcomes, summative outcomes, perceived purposes of appraisal, overall effectiveness of appraisal, and summative purposes such as promotion and dismissal of staff. Principal components analysis and confirmatory analysis yielded the two a priori outcome factors, each of which was significantly correlated with perceived overall effectiveness of appraisal. When analyzed separately, analysis of variance found that senior teachers appraised by the school principal (SP) had a stronger sense of the formative purpose of appraisal than teachers appraised by senior staff (TS) whereas the TS group had the weakest sense of the promotion purposes. Teachers appraised by the principal (TP) had a stronger perception of the dismissal purpose than the TS group. The TS group was more favorable toward the competence of their appraisers than the TP group. Although the three groups did not differ in their perceptions in formative outcomes, summative outcomes, or overall effectiveness of teacher appraisal, the appraiser-appraisee combination did make a significant difference in teacher perceptions of the purposes and appropriateness of the appraisal.

Staff appraisal in the school has developed along summative and formative lines. Summative appraisal often forms the basis for initial certification of teachers, renewal of contracts, and perhaps promotion and dismissal of staff (Parker, 1985; Turner & Clift, 1988; Wells, 1989). Formative appraisal emphasizes personal development through the appraisal process (Castetter, 1992). In terms of the effectiveness of teacher appraisal, therefore, the major concern of summative appraisal is managerial decision whereas the major concern of formative appraisal is the provision of information for the improvement of teachers and their teaching. The present investigation extended Mo, Conners, and McCormick's (1998) study which hypothesized that teachers' perceived effectiveness of staff appraisal is associated both with their perceptions of the formative and summative outcomes of the appraisal process. We applied confirmatory factor analysis to provide a stronger validation of these outcome constructs. We then examined whether teacher perceptions of staff appraisal would vary when teachers of lower and higher ranks are appraised by an appraiser of varying power and authority.

Staff Appraisal in Hong Kong

In its seventh report, the Education Commission (EC, 1997) confirms to support the School Management Initiative (EMB & ED, 1991)--now under the broader term of school-based management (SBM)--as one of the means to achieve internal quality assurance. The Education Commission Report No. 7 (ECR7) argued that quality assurance within schools can be best achieved through implementing school-based management, allowing key players of school education to participate in setting school goals and developing quality indicators.

For historical reasons, Hong Kong has been greatly influenced by the British system of education. Supervision and appraisal of teachers have long been the responsibility of the Advisory Inspectorate Division of the Education Department (ED). The inspectorate carries out inspections and visits to schools, the major purpose of which is to monitor and improve the quality of teaching (Education Department, 1992a). The evaluation of school performance greatly depends on the effectiveness of the teachers' teaching in their classrooms. Therefore a major task of the inspectorate is to come to schools and do random checking mainly through classroom observations. Because of the huge amount of manpower involved in full inspection to around 1,200 primary and secondary schools, not all teachers can be assessed in each inspection cycle. In some extreme cases, a full inspection cycle for a certain subject area might need a duration as long as fifteen years (Education Commission, 1994). Apart from general school inspections, teachers might also be appraised formally by representatives from ED upon request for specific purposes. Because schools in Hong Kong are highly structured with their career ladder system, and teachers are paid according to their ranks (Education Department, 1994), a teacher will be formally appraised by an inspector of the Advisory Inspectorate Division if he or she is nominated for promotion.

However, with the full implementation of SBM after September 2000, schools are given the authority to go over the promotion assessment exercise on their own. For purposes of selection for promotion to a higher rank, those teachers will normally be appraised by their superordinates. In the case of primary schools, usually the school principal will take the role as the appraiser. In the secondary, sometimes the senior teacher will be appointed by the school head as appraiser to carry out the appraisal procedure and report to the principal for recommendation.

Before the release of the School Management Initiative (SMI) in 1991, formal school-based teacher appraisal was practiced in only a few schools in Hong Kong (Cheng, 1992). However, since the advent of the SMI, those schools which have joined the scheme have to implement a staff appraisal system to evaluate the performance of their teachers. According to the Education Department (1992b), teacher appraisal in Hong Kong schools participating in the SMI scheme is expected to serve both formative and summative purposes, including helping teachers improve their performances through appropriate staff development, identifying the potential of teachers for career development, and

providing information for making decisions about staff promotion and disciplinary actions.

However, even after most schools have joined the SBM since September 2000, some schools are still reluctant to change their existing management systems. Recognizing that there should be no more delay, the Education Department finally set a time frame for implementing various elements of school-based management. In particular, schools are required to implement an appraisal system before the end of the 2001-02 school year (Education Department, 1999). There is therefore an urgent need for establishing an effective staff appraisal system.

Staff Appraisal and its Purposes

A well-planned and carefully implemented teacher appraisal system can have a far-reaching impact on teacher effectiveness, while a poorly planned one can dampen staff morale and have a negative effect on teacher performance (Larson, 1984). Good practices of staff appraisal system may enable personal and professional growth of teachers and form the basis for fair and constructive personal decisions. To Stronge's (1991) perception, if an appraisal system does not have a clear purpose, it will just be a meaningless exercise.

Researchers have distinguished between the formative and summative purposes of staff appraisal. Whereas Valentine (1992) has noted that purely formative systems do not address the need to make personnel decisions based on the competency of teachers, other researchers, like Schalock & Schalock (1993) and Wright, Horn & Sanders (1997), argued that the concern for accountability for improved student learning and the notion that learning gains by student is the most important accomplishment of teachers have led to the development of summative systems. However, an interesting issue which has received little attention is the potentially varying effects of different appraisers on the appraisee, whether the appraisal be formative or summative in nature.

Who Appraises Whom and What Really Matters?

With reference to Mo, Conners & McCormick (1998), appraisers are important as they can influence the appraisees directly. Therefore, in the school context, the relationship between teacher and appraiser is central to successful outcomes of appraisal (Duckett, 1991). If the appraisal is done by someone in a line management position, it is important that the appraiser be credible, respected, and skillful in appraising teachers so as to eliminate the fear of misuse of appraisal data (McNamara, 1995). For suggestions to be accepted by the teachers for facilitating their growth, the appraiser should be helpful, patient, trustworthy, credible in providing useful information, able to demonstrate new ideas and techniques, and able to persuade teachers with convincing reasons (Duke & Stiggins, 1986). These characteristics of the appraiser is particular important for formative purposes because the ultimate goal of this kind of appraisal is the improvement of teacher performance. The teacher, being the appraisee, may then improve if the appraiser is competent in providing appropriate feedback on any problems

identified (Natriello, 1990).

The Present Study

The purposes of the present investigation were: (a) to validate the formative and summative outcome measures and examine their relations with perceptions of overall effectiveness of appraisal as suggested by Mo et al. (1998), and (b) to examine potential differences of perceptions of teachers appraised by their senior teacher or by their school principal, and senior teachers appraised by their principal.

Method

The Sample

The participants were 550 teachers from 20 elementary schools in Hong Kong. Written consent was obtained from the teachers before they completed the survey. After listwise deletion of missing data, the analysis used a sample of 527 teachers who had complete data for confirmatory factor analysis (CFA). Of the 527 teachers, 201 were teachers appraised by senior teachers in their school, 269 were appraised by their school principals, and 69 were senior teachers appraised by their principals. In subsequent analysis of variance, the responses of these three groups of teachers were compared. Teachers whose rank or appraiser were unidentified were excluded from the analysis.

Measures

Based on Mo et al. (1998), the instrument consisted of 26 items forming seven measures (see Appendix). The two staff appraisal outcomes included formative and summative appraisal outcomes derived from 10 and 4 items, respectively. A formative appraisal purpose construct was derived from 3 items. The summative appraisal purposes were a promotion and a dismissal item. Also included was an appraiser competence (6 items) and an overall perception of appraisal effectiveness measure. They are described below:

Formative outcome. Ten items asked the teachers about their perceived outcomes of teacher appraisal in terms of potential improvement in teaching and professional teacher development. Two items that were used in the original Mo et al. (1998) instrument were excluded because of their potential cross loadings on both the formative and summative outcome constructs (see Mo et al., 1998).

Summative outcome. Four items asked the teachers about their perceived summative outcomes of teacher appraisal such as managerial decisions on employment and career advancement.

Formative appraisal purpose. Three items asked the teachers their perception of the formative purpose of teacher appraisal in improving teaching.

Appraiser competence. Six items asked teachers how competent they believed their appraiser was in the appraisal process.

Promotion. A single item asked teachers how they agreed to the summative purpose of appraisal for promotion to a higher rank.

Dismissal. A single item asked teachers how they agreed to the summative purpose of appraisal for the managerial decision of dismissal of staff. The promotion and dismissal purposes were analyzed separately because Mo et al. (1998) found a low reliability of the summative purpose construct combining both these purposes.

Overall effectiveness. A single item asked teachers about their perceptions of the overall effectiveness of staff appraisal in their schools.

Insert Figure 1 About Here

Statistical Analyses

In preliminary analysis, we first examined the reliability of each a priori construct with multiple indicators. Consistent with Mo et al. (1998), we conducted a principal components analysis to test the ability of 10 items to form a formative outcome factor and 4 items to form a summative outcome factor. To extend previous findings, we followed up with a confirmatory factor analysis (CFA) to test the factor structure of these two a priori factors. To further validate the two appraisal outcome factors, we added to the model a formative purpose factor derived from three items to examine its relation with the outcome factors. Based on the CFA factors with multiple indicators, we then conducted multivariate analysis of variance (ANOVA) on the scale means of the three dependent variables (formative outcome, summative outcome, and formative purpose) and on four single-item variables (promotion purpose, dismissal purpose, appraiser competence, and overall perception of appraisal effectiveness).

The conduct of SEM has been described elsewhere (e.g., Bentler, 1990; Bollen, 1989; Byrne, 1998; Joreskog & Sorborm, 1993; Marsh, Balla, & Hau, 1996; Pedhazur & Schmelkin, 1991) and is not further detailed here. All analyses throughout this paper were conducted with the SPSS version of PRELIS and LISREL using maximum likelihood procedures (Joreskog & Sorbom, 1988). The goodness of fit of models is evaluated based on suggestions of Marsh, Balla, and McDonald (1988) and Marsh, Balla, and Hau (1996) with an emphasis on the Tucker-Lewis index (TLI), but we present also the chi-square test statistic and the relative noncentrality index (RNI). The CFA models were based on a 24 x 24 covariance matrix for the 4 scales (10 items for formative outcome, 4 items for summative outcome, 3 items for formative purpose, 1 item for overall effectiveness, and 6 items for appraiser competence).

CFA models tested. To first validate the formative and summative outcome measures, in models 1 and 2 (Figure 1) we tested whether a 2-factor structure fitted better than a single outcome factor model (models 3 and 4). Model 1 differed from model 2 in that model 2 had two correlated

uniquenesses (correlations between the residuals of response items) included in the model. Similar correlated uniquenesses were included in model 4 (see Appendix). Support for a 2-factor structure requires model 1 to fit better than model 3 (without correlated uniquenesses) and model 2 to fit better than model 4 (with correlated uniquenesses included). When the 2-factor structure was established, we added a formative purpose construct derived from 3 items (Model 5). Strong validation of the 2-factor structure requires the formative purpose factor to correlate higher with the formative outcome than with the summative outcome factor. This application using an external criterion variable to validate the CFA factor structure is well documented (see Yeung, Chui, Lau, McInerney, Russell-Bowie, & Suliman, 2000; Yeung & Lee, 1999). Model 5 also provided an appropriate validation for the formative purpose measure for subsequent analysis. In model 6, instead of the formative purpose measure, we added to model 2 a single-item overall effectiveness measure to test its relation to each of the outcome factors. Consistent with Mo et al. (1998), we expected that overall effectiveness would be correlated significantly with both the formative and summative outcomes. Model 7, based on model 5, included the six items for appraiser competence to establish the validity of this construct for subsequent analysis.

Analysis of variance. The mean of the four scales with multiple indicators were obtained by averaging the item scores in each respective construct. Using multivariate ANOVA, a total of seven dependent variables (formative outcome, summative outcome, formative purpose, promotion, dismissal, appraiser competence, and overall effectiveness) were compared across three groups (teachers appraised by senior teachers, teachers appraised by principals, senior teachers appraised by principals). Variables that were found to differ across groups were further analyzed with a oneway ANOVA followed by Scheffe range tests to determine which group had more favorable perceptions in each of these variables.

Results

Preliminary Analysis

The alpha reliability estimates (see Appendix) for the four scales were good (αs from .72 to .93). Consistent with Mo et al. (1998), although presumably pertaining to a summative purpose construct, the Promotion and Dismissal measures had an unreasonably low reliability ($\alpha < .5$). Therefore in the analysis, we treated Promotion and Dismissal as two distinct measures.

Principal components analysis of the 14 items of teacher appraisal outcomes with varimax rotation (Norusis, 1994) revealed the two a priori factors with eigenvalues 7.56 and 1.16 respectively that together accounted for 62.3% of the total variance. The factor coefficients were good (from .63 to .78 for the Formative Outcome factor and from .69 to .79 for the Summative Outcome factor).

Insert Table 1 About Here

CFA Models

A summary of the goodness of fit of CFA models is presented in Table 1. Model 1 positing two outcome factors (Figure 1) provided a better fit than model 3 positing a single outcome factor (TLI of .89 vs. .85). Similarly, model 2 positing two factors provided a better fit than model 4 positing a single factor (TLI of .94 vs. .89). Because model 2 with correlated uniquenesses provided a good fit to the data (TLI > .9), subsequent models had these correlated uniquenesses included. The factor coefficients of models 1 and 2 were good (.66 to .80). The correlation between the formative and summative outcomes were reasonably high (r = .79). This result supported a 2-factor outcome structure but also that the two outcomes were quite highly associated.

Model 5 included three items for the formative purpose construct used as an external criterion to test the validity of the outcome variables. Model 5 had a good fit (TLI = .93). The factor coefficients were good (.61 to .81). Strong validation of the two-outcome structure would require formative purpose to correlate higher with the formative outcome than with the summative outcome construct. An inspection of the factor correlations found that formative purpose correlated with formative outcome at .45, higher than the correlation with summative outcome (.39). Whereas this result provided some support for the validity of the formative-summative outcome distinction, the moderate correlation between Formative Purpose and Summative Outcome (r = .39) together with the high correlation between the two Outcome measures (r = .79) also indicated a surprisingly close relation between these Outcome constructs.

Model 6 tested the hypothesis that both outcomes were related to an overall effectiveness of staff appraisal. Model 6 provided a good fit (TLI = .92). The parameter estimates are presented in Table 2. The correlation of Overall Effectiveness with Formative Outcome (r = .64) and with Summative Outcome (r = .65) were both significant, indicating that both outcomes were related to perceptions of overall effectiveness of staff appraisal.

Model 7 included Formative Outcome, Summative Outcome, Formative Purpose, and Appraiser Competence to establish the validity of these constructs for subsequent analysis of variance. Model 7 provided a good fit to the data (TLI = .93). The factor coefficients were good (.58 to .80) and the factor correlations were mostly moderate (rs from .39 to .79). These results provided good support for the distinctiveness of the four constructs with multiple indicators. Thus model 7 provided a good basis for subsequent comparison of these variables across three groups of teaching staff.

Insert Tables 2 and 3 About Here

Multivariate ANOVA

Seven dependent variables were compared across three groups of teaching staff. The scores of four variables (formative outcome, summative outcome, formative purpose, and appraiser competence) were derived from the average of items for each respective scale. The other three variables (promotion, dismissal, and overall effectiveness) were from scores of single items. The mean and standard deviation of each variable is presented in Table 3.

Multivariate ANOVA results found significant differences of scores across groups. The univariate F-statistics are shown in Table 3. Of the seven variables, four were found to differ across groups. Oneway ANOVA followed by Scheffe range tests found that compared to senior teachers appraised by their school principals, teachers who were assessed by senior teachers in their schools had weaker perceptions of the formative purpose of staff appraisal. Teachers appraised by senior teachers also had the weakest perception of the summative purpose of promotion among the three groups. They had weaker perceptions of the summative purpose of dismissal but stronger perceptions of their appraiser's competence in the appraisal process than teachers appraised by their school principals.

Discussion

An important issue in teacher appraisal is the validity of a formative and summative measure. The analysis in the present study found that these are two distinct constructs, each relating to overall teaching effectiveness. The present CFA models have shown a useful approach to construct validation by including some external criteria to the models. The present CFA results are consistent with Mo et al (1998) that used an exploratory factor analysis approach. The present CFA approach has, however, provided us with not only a stronger validation of the scales for subsequent inter-group comparisons, but also a better understanding of the relationships among the constructs considered here. In particular, there is strong support for the distinction between the formative and summative outcomes of staff appraisal.

In comparing among groups of teachers, it is interesting to learn that the senior teacher respondents seemed to be more inclined to recognize the "expertise" (whether due to their authority power or their subject knowledge expertise) of their "boss" in appraising them for formative purposes of appraisal. Perhaps, the senior teachers could more readily accept their boss to be their appraiser because they have had close contacts with their school head and they are perhaps more familiar with their leadership style and the structural and bureaucratic characteristics of the school as a formal organization. But to the teacher subordinates, they may have reservations when appraised either by their senior members or their "super boss", the principal. Perhaps this is actually reflecting the fact that most teachers, especially in the primary schools of Hong Kong, are very conservative and they have much confidence in their teaching performances. They probably do not want to be disturbed in their classroom 'empire', especially when they are forced to be observed. But if they are forced to choose in between the

two appraisers, they would choose to be observed by their senior colleagues whom they believe are more competent in providing considerably better advice to them, especially for formative purposes. Therefore the present study found that responses of teachers appraised by their senior staff (group 1 in Table 3, $\underline{M}$ = 3.28) were slightly more favourable toward the competence of their appraisers than teachers appraised by their principal (group 2 in Table 3, $\underline{M}$ = 3.14). That is perhaps why staff appraisal through classroom observation is still not commonly accepted by most school teachers of Hong Kong. They regard themselves as experienced professionals who are expert enough to carry out their routine teaching duties. With reference to the normal promotion practices in schools, if teachers are formally appraised by their senior members instead of their principal, conflicts or problems might occur.

Moreover, teachers appraised by their principal (group 2 in Table 3, $\underline{M}$ = 2.93) had a comparatively stronger perception of the dismissal purpose than teachers appraised by their senior teachers (group 1 in Table 3, $\underline{M}$ = 2.67) and senior teachers appraised by their principal (group 3 in Table 3, $\underline{M}$ = 2.67). However, it is worth noting that all the means are below the mid-point 3. In other words, the respondents did not feel comfortable to be appraised by their superordinates, whatever positions they were. Thus, if teachers are to be appraised for potential dismissal purpose, they would prefer not to be appraised by their line managers (i.e., the senior teachers), who are supposed to have no official authority to dismiss any staff. Rather, for potential dismissal decisions, they would choose to have their principal, whom they believe is their official appraiser equipped with formal authorization by both the government and the respective school management board of the school sponsoring body concerned.

In sum, although the three groups did not differ in their perceptions in formative outcomes, summative outcomes, or overall effectiveness of teacher appraisal, the appraiser-appraisee combination did make a significant difference in teacher perceptions of the purposes and appropriateness of the appraisal. Consistent with Mo et al. (1998), the finding in the present study indicates that the overall perceived effectiveness of a staff appraisal system was related to both its formative and summative outcomes. An examination of the potential differences of teacher perceptions found that senior teachers appraised by their principal had a stronger positive sense of the formative purpose of appraisal than teachers appraised by senior teachers. However, for summative purposes, teachers tended to prefer their principal to do the appraisal for them.

References

Bentler, P. M. (1990). Comparative fit indices in structural models. Psychological Bulletin, 107, 238-246.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Byrne, B. M. (1998). Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming. Mahwah, NJ: Erlbaum.

Castetter, W. B. (1992). The personnel function in educational administration (5th ed). New York: Macmillan.

Cheng, Y.C. (1992). A preliminary study of School Management Initiative: Responses to induction and implementation of management reforms. Educational Research Journal, 7, 21-32.

Duckett, I. (1991). Promoting appraisal through an active staff development programme. School Organization, 11 (2), 153-169.

Duke, D. L. & Stiggins, R. J. (1986). Teacher evaluation: Five keys to growth. Washington, D. C.: National Educational Association.

Education and Manpower Branch and Education Department (EMB & ED) (1991). The School Management Initiative: Setting the Framework for Quality in Hong Kong Schools. Hong Kong: Government Printer.

Education Commission. (1994). Report of the working group on educational standards: Quality in school education. Hong Kong: Government Printer.

Education Commission. (1997). Education Commission Report No. 7: Quality school education. Hong Kong: Government Printer.

Education Department. (1992a). SMI manual on school administration. Hong Kong: Education Department.

Education Department. (1992b). Staff appraisal in schools. Hong Kong: Education Department.

Education Department. (1994). Code of Aid. Hong Kong: Education Department.

Education Department. (1999). Implementation of school-based management in all public sector schools. Administration Circular No. 3/99.

Joreskog, K. G., & Sorbom, D. (1988). LISREL 7: A guide to the program and applications. Chicago: SPSS.

Joreskog, K. G., & Sorbom, D. (1993). LISREL 8: Structural equation modeling with SIMPLIS command language. Chicago: Scientific Software International.

Larson, (1984). Teacher performance evaluation: What are the key elements? *NASSP Bulletin, 68* (466), 13-18.

Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical process. In G. A. Marcoulides & R. E. Schumacker (Eds.), Advanced structural equation modeling techniques (pp. 315-353). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in comfirmatory factor analyses: the effect of sample size. Psychological Bulletin, 103, 391-410.

McNamara, N. (1995). Teacher appraisal. Sydney: Catholic Education Office.

Mo, K. W., Conners, R., & McCormick, J. (1998). Teacher appraisal in Hong Kong self-managing secondary schools: Factors for effective practices. Journal of Personnel Evaluation in Education, 12, 19-42.

Natriello, G. (1990). Intended and unintended consequences: Purposes and effects of teacher evaluation. In J. Millman and L. Darling-Hammond (Eds.), The new handbook of teacher evaluation: Assessing elementary and secondary school teachers (pp. 35 - 45) Newbury Park, CA: Corwin Press.

Norusis, M.J. (1994). SPSS professional statistics 6.1. Chicago, IL: SPSS.

Parker, J. C. (1985). Career ladder/master teacher programs: Implications for principals. Reston, VA: The national Association of Secondary School Principals.

Pedhauzur, E. J., & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

Schalock, H. D. & Schalock, M. D. (1993). Student learning in teacher evaluation and school improvement: An introduction. Journal of Personnel Evaluation in Education, 7 (2), 103-104.

Stronge, J. H. (1991). The dynamics of effective peformance evaluation systems in education: Conceptual, Human relations, and technical domains. Journal of Personnel Evaluation in Education, 5 (1), 77-83.

Turner, G., & Clift, P. (1988). Studies in teacher appraisal. East Sussex, UK: Falmer Press.

Valentine, J. W. (1992). Principles and practices for effective teacher evaluation. Needham Heights, MA: Allyn and Bacon.

Wells, P. (1989). Policy developments in the United States of America. In J. Lokan & P. McKenzie (Eds.), Teacher appraisal: Issues and practices. Radford House, Victoria: ACER.

Wright, S. P., Horn, S. P. & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications fro teacher evaluation. Elementary School Journal, 11 (1), 57-67.

Yeung, A. S., Chui, H. S., Lau, I. C. Y., McInerney, D. M., Russell-Bowie, D., & Suliman, R. (2000). Where is the hierarchy of academic self-concept? Journal of Educational Psychology, 92, 556-567.

Yeung, A. S., & Lee, F. L. (1999). Self-concept of high school students in China: Confirmatory factor analysis of longitudinal data. Educational and Psychological Measurement, 59, 431-450.

Table 1.

Goodness-of-fit Summary for Models

Model	χ^2	df	TLI	RNI	Null χ^2	df	F.L.
1. 2 Outcome factors, no CU	459.62	76	.898	.915	4605.35	91	.66-.80
2. 2 Outcome factors, 2 CU	292.64	74	.940	.952	4605.35	91	.66-.80
3. 1 Outcome factor, no CU	638.83	77	.853	.876	4605.35	91	.57-.80
4. 1 Outcome factor, 2 CU	463.84	75	.895	.914	4605.35	91	.58-.80
5. 2 Outcomes + Formative Purpose	379.91	114	.936	.947	5111.96	136	.61-.81
6. 2 Outcomes + Overall Effectiveness	389.86	86	.924	.938	5000.35	105	.67-.80
7. Model 5 + Appraiser Competence	603.99	222	.932	.940	6615.94	253	.58-.80

Note: RNI= Relative noncentrality index. TLI= Tucker-Lewis index. CU = correlated uniquenesses. N = 539.

Table 2.

Solution of Model 6

Variables	Formative Outcome		Summative Outcome		Overall Effectiveness	
	Factor Coeff	Unique	Factor Coeff	Unique	Factor Coeff	Unique
Item 1	.70*	.52*	.69*	.52*	1	0
Item 2	.69*	.52*	.70*	.51*	--	--
Item 3	.73*	.46*	.73*	.46*	--	--
Item 4	.77*	.41*	.74*	.45*	--	--
Item 5	.75*	.43*	--	--	--	--
Item 6	.80*	.36*	--	--	--	--
Item 7	.79*	.37*	--	--	--	--
Item 8	.66*	.56*	--	--	--	--
Item 9	.79*	.38*	--	--	--	--
Item 10	.80*	.37*	--	--	--	--

Factor Correlations

Formative Outcome	--		
Summative Outcome	.79*	--	
Effectiveness	.64*	.65*	--

Note: N= 539. Parameters estimates are completely standardized. *p < .05.

BEST COPY AVAILABLE

Table 3.

Means and (Standard Deviations) of Variables in 3 Groups

Variables		A (T by ST)	B (T by P)	C (ST by P)	Univ. F	MSE	Scheffe
		n = 201	n = 269	n = 69	(2,536 df)		
Formative outcome	M	3.03	3.07	3.03	0.32	0.45	--
	SD	(0.66)	(0.67)	(0.73)			
Summative outcome	M	2.92	2.96	2.96	0.23	0.23	--
	SD	(0.63)	(0.70)	(0.70)			
Formative purpose	M	3.63	3.66	3.85	3.60*	0.36	C>A
	SD	(0.62)	(0.61)	(0.53)			
Promotion	M	3.02	3.30	3.35	6.30*	0.81	C,B>A
	SD	(0.96)	(0.86)	(0.89)			
Dismissal	M	2.67	2.93	2.67	4.46*	1.01	B>A
	SD	(1.03)	(0.95)	(1.13)			
Appraiser competence	M	3.28	3.14	3.18	3.41*	0.34	A>B
	SD	(0.57)	(0.57)	(0.64)			
Overall Effectiveness	M	2.89	2.95	2.88	0.44	0.61	--
	SD	(0.80)	(0.75)	(0.83)			

Note: Group A (T by ST) are teachers appraised by senior teachers. Group B (T by P) are teachers appraised by school principals. Group C (ST by P) are senior teachers appraised by school principals. N = 539. * $p < .05$

BEST COPY AVAILABLE

Appendix

Items Used in the Present Study

Formative Outcome Alpha = .93

1. There has been an increase in my teaching skill since my last appraisal. [a]
2. The appraisal makes me reflect more on my teaching practices. [a]
3. As a result of my last appraisal I now consider teaching more like a professional job. [b]
4. The appraisal makes me know the direction of my professional development. [b]
5. As a result of my last appraisal I have improved my class management.
6. I have a better understanding of the teaching-learning process since my last appraisal.
7. Because of the appraisal I have a better understanding of my strengths and weaknesses.
8. Teacher appraisal makes me care more about teaching.
9. Because of the appraisal I gain more reinforcement on my teaching.
10. Teachers in my school understand their strengths and weaknesses because of the appraisal.

Summative Outcome Alpha = .81

1. Teacher appraisal in this school increases the accountability of the school.
2. Teacher appraisal in this school provides useful information for the school to make managerial decisions.
3. Teacher appraisal in this school enables the fair promotion of staff.
4. Teacher appraisal in this school provides a fair assessment of the performance of teachers.

Competence of Appraiser Alpha = .82

1. The appraisal was trustworthy during the appraisal process.
2. The appraisal was openminded during the appraisal process.
3. The appraiser had a good interpersonal relationship with me.
4. The appraiser was skillful in teaching.
5. The appraiser was familiar with the subject matter that I taught.
6. The appraiser had considerable experienced in teaching.

Formative Purposes of Appraisal Alpha = .72

1. One of the purposes of teacher appraisal is to enhance teacher reflection on their teaching practices.
2. A purpose of teacher appraisal is to help teachers know their strengths and weaknesses and areas that require improvement.
3. Finding out the professional development needs of teachers is one of the purposes of teacher appraisal.

Promotion

1. One of the purposes of teacher appraisal is promotion of staff.

Dismissal

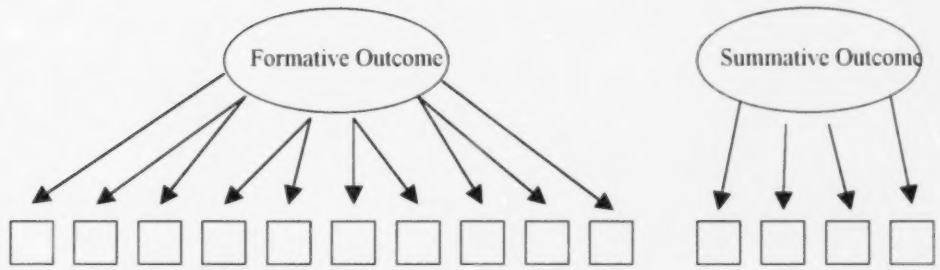
1. Gathering information for teacher dismissal is one of the purposes of teacher appraisal.

Overall Effectiveness

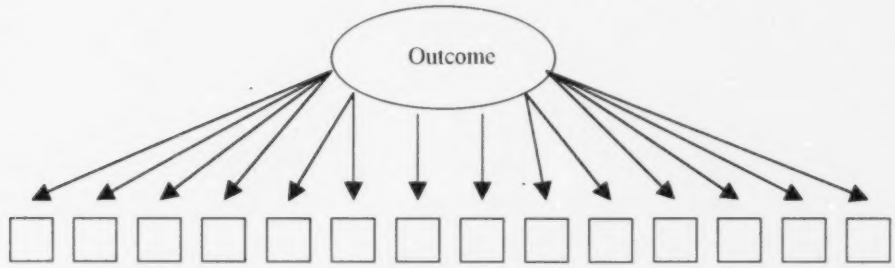
1. Overall, the appraisal system of this school is effective.

Note: [a b] The uniquenesses of items with the same superscript were correlated in CFA models.

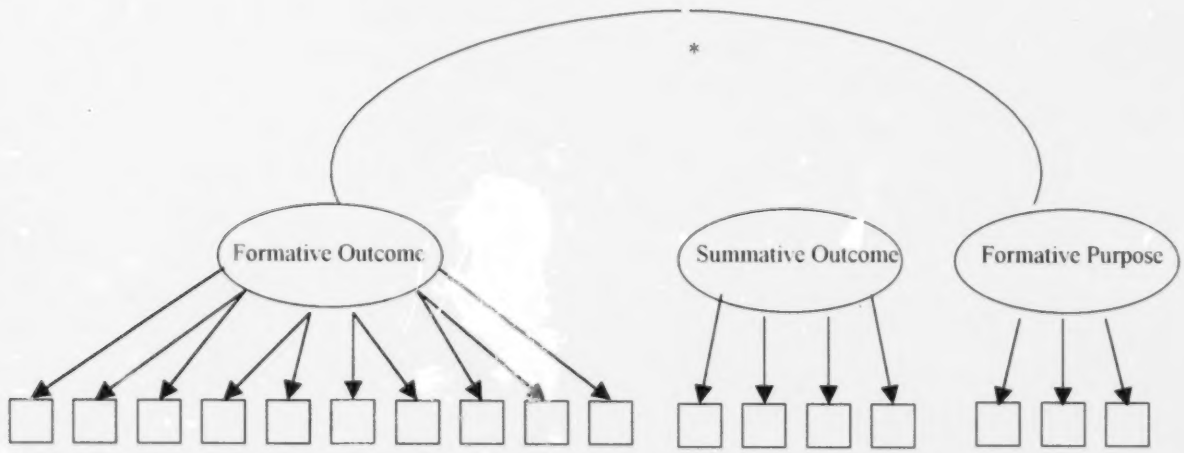
Figure 1. Confirmatory factor analysis models.



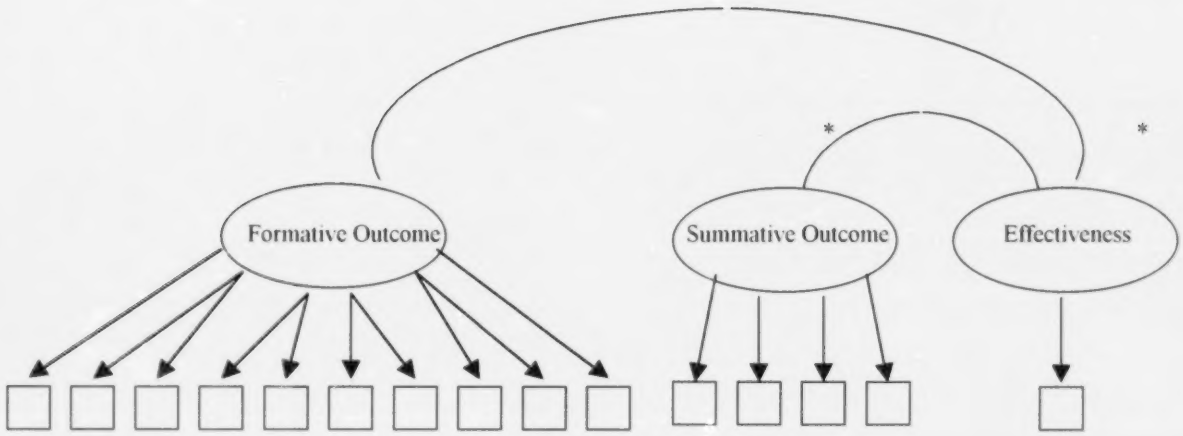
(a) Models 1 and 2 positing 2 Outcome factors.



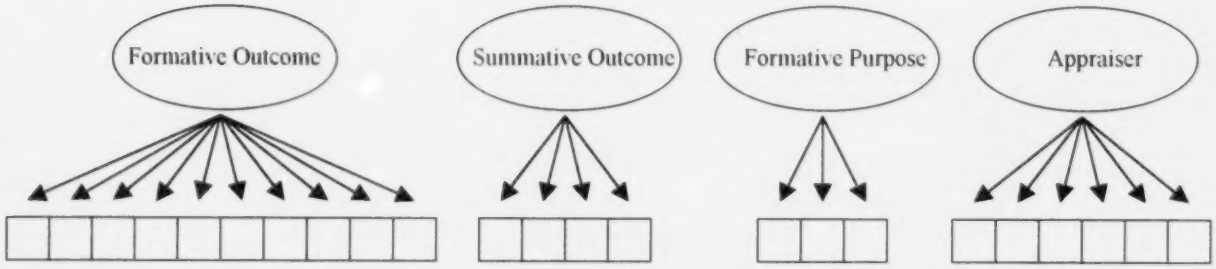
(b) Models 3 and 4 positing 1 Outcome factor.



(c) Model 5 positing 2 Outcome factor and 1 Formative Purpose factor.



(d) Model 6 testing the correlations between Overall Effectiveness and Outcome factors.



(e) Model 7 positing 2 Outcome factor, 1 Formative Purpose factor, and 1 Appraiser Competence factor.